Interlaboratory Comparison Measurement EstOil-8

Final Report

11.02.2014

Report Compiled by

Lauri Jalukse

Ivo Leito

Tiina Kukk

This report is available at the website of UT at http://www.ut.ee/katsekoda/ILC/

Testing Centre, University of Tartu

Scanola Baltic Ltd

Tartu 2014

Table of Contents

1 The Aim of the Intercomparison

The aim of the EstOil-8 intercomparison was to enable the participating laboratories to assess their performance in determining seven edible oil parameters: moisture content, free fatty acids content (below FFA), peroxide value (below PV), saponification value (below SAPV), beta-sitosterol content (below B-SITO), phosphorus content (below P) and erucic acid content (below EA). All parameters except P were determined in refined rapeseed oil (P was determined in crude rapeseed oil). The previous intercomparisons: EstOil-1[1], EstOil-2[2], EstOil-3[3], EstOil-4[4], EstOil-5[5] , EstOil-6[6] and EstOil-7[7] took place in 2005, 2006, 2007, 2008, 2009, 2010 and 2011 respectively.

2 Organization

2.1 General

The intercomparison measurement was organized jointly by the Testing Centre of University of Tartu (below UT) and Scanola Baltic Ltd (below SB). See Table 1 for the detailed contact information of the organizers.

Table 1. Contact Information of the Organizers.

University of Tartu, Testing Centre
http://www.ut.ee/katsekoda
Ravila 14a, 50411 Tartu, Estonia
Phone: +372 51 84 176
Fax: +372 737 5264
E-mail: lauri.jalukse@ut.ee

Scanola Baltic Ltd
http://www.werol.ee
Painküla 48422, Jõgevamaa, Painküla, Estonia
Phone: +372 77 68233
Fax: +372 77 68 220
E-mail: tiina.kukk@olivia.eu

This report was compiled jointly by UT and SB and is publicly available via the website of UT at http://www.ut.ee/katsekoda/ILC/. The participants are listed in this report but the results are presented in random order, so that the results cannot be traced back to the participants. Every participant will receive a private letter revealing his/her result number and permitting assessment of performance.

2.2 The Samples

The oil samples were prepared and distributed by SB. The samples were refined rapeseed oil[8] (moisture content, FFA content, PV, SAPV, B-SITO content, EA content) and crude rapeseed oil (P content) samples of approximately 100 ml in gas-tight (sealed) amber glass bottles. The samples were prepared from a single bulk of oil that was well mixed before filling the bottles. The bottles were filled and closed during a short time (around 30 seconds per bottle). The laboratories got random bottles from the pool of bottles. The first and last bottles were not distributed.

[1] http://www.ut.ee/katsekoda/ILC/Estoil/Estoil_1_rep_Final.pdf

[2] http://www.ut.ee/katsekoda/ILC/Estoil/EstOil-2_2006_final_report.pdf

[3] http://www.ut.ee/katsekoda/ILC/Estoil/EstOil-3_final_report.pdf

[4] http://www.ut.ee/katsekoda/ILC/Estoil/EstOil-4_Final_Report.pdf

[5] http://www.ut.ee/katsekoda/ILC/Estoil/EstOil-5_Final_Report.pdf

[6] http://www.ut.ee/katsekoda/ILC/Estoil/EstOil-6_Final_Report.pdf

[7] http://www.ut.ee/katsekoda/ILC/Estoil/EstOil-7_Final_Report.pdf

[8] Antioxidant (vitamin e) was added to the PV samples of refined rapeseed oil for better stability of peroxide value.

2.3 Data Treatment

The evaluation of participant data was done at UT according to the EN ISO-IEC 17043:2010[9] and standard ISO 13528:2005.[10] The z-score approach was used. The z-score for a particular measured value of a participant is calculated according to the following equation:

$$z = \frac{x - x_c}{s}, \quad (1)$$

where x is the participant's value, x_c is the consensus value and s is the target standard deviation. The consensus values and target standard deviations were found using the Algorithm A described in the ISO 13528:2005 standard[10]. This algorithm gives the so-called robust estimates of the consensus value and standard deviation of participants.

The moisture content was specified in the invitations to EstOil-8 as to be determined by the Karl Fischer procedure. Therefore we did not include the results of gravimetric (heating loss) moisture determination into the data set that was used for calculation of the consensus value and target standard deviation. Nevertheless, such results were retained for assessment of participant performance.

Assessment of participant performance was carried out in two ways.

(1) Absolute values of z-scores (|z| values) are used for assessing the acceptability of the results as described in Table 2.

Table 2. Assessment of Acceptability of the Results Using z-Scores.

|z| Value	Acceptability of the Result	Required Action
$\|z\| \leq 2$	Acceptable result	No action is required
$2 < \|z\| < 3$	Doubtful result	Preventive action is required
$\|z\| \geq 3$	Unacceptable result	Corrective action is required

(2) Pairwise E_n values between participants presented as tables.

This is done using the E_n numbers as described in EN ISO-IEC 17043:2010:[9]

$$E_n = \frac{C_{lab1} - C_{lab2}}{\sqrt{U_{lab1}^2 + U_{lab2}^2}} \ . \quad (2)$$

where C_{lab1} and C_{lab2} are the results of the two laboratories that are compared and U_{lab1} and U_{lab2} are their expanded uncertainties. Equation 2 is adequate, if between-sample variability is significantly (more than 5 times) lower than (between-participants) target standard deviation. If not, the between-sample variability has to be taken into account and the E_n value is found as follows:

$$E_n = \frac{C_{lab1} - C_{lab2}}{\sqrt{U_{lab1}^2 + U_{lab2}^2 + (t_{95}(df) \cdot s_s)^2}} \ . \quad (3)$$

where s_s – is the between-sample standard deviation and $t_{95}(df)$ is the student coefficient at 95% confidence level with df degrees of freedom. Agreement between two results is considered

[9] EN ISO-IEC 17043:2010 *Conformity assessment – General requirements for proficiency testing*, ISO/IEC 2010 (This standard replaces the ISO Guides 43-1 and 43-2).

[10] ISO 13528:2005. *Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparisons*, ISO, 2005.

acceptable if $|E_n| \leq 1$. Participants who did not report uncertainties for their results were excluded from the pair-wise comparisons.

In addition to the above-described data treatment schemes we similarly to EstOil-3, EstOil-4, EstOil-5, EstOil-6 and EstOil-7 carried out data treatment according to the "robust statistics" approach[11], which is presented in Annex 1. This approach permits avoiding some of the problems of the two standard approaches presented above. Since this approach is not included in the leading international standards on interlaboratory comparisons the results obtained with it are informatory only.

3 Participants

Invitations were sent to a number of European laboratories. The participants are listed in Table 3.

Table 3. Participants to EstOil-8.

Institution	Country
Herkon d.o.o.	Bosnia
CONTROL-H d.o.o. Mostar	Bosnia
Laboratory quality control Zvijezda	Croatia
Scanola Baltic Ltd	Estonia
Agricultural Research Centre	Estonia
Central Cemistry Laboratory of the Health Board_Estonia	Estonia
Tartu Veterinary and Food Laboratory	Estonia
Testing Centre, University of Tartu	Estonia
Institute of Ecology and Earth Sciences, University of Tartu	Estonia
Union of Agricultural Cooperatives of Iraklion	Greece
Galanakis Laboratories	Greece
National Food Chain Safety Office Reg. Food Chain Lab Miskola	Hungary
National Food Chain Safety Office, Food Analytical National Reference Laboratory of Food and Feed Safety Directorate	Hungary
AS "Latvenergo" Chemical laboratory	Latvia
Institut za hmeljarstvo in pivovarstvo Slovenije Slovenia	Slovenia

[11] Jörg W. Müller, *J. Res. Natl. Inst. Stand. Technol.* **2000**, 105, 551-555

4 Results

4.1 Results of the Participants

Results of the participants are presented in Table 4. The results are presented with the same number of decimal digits as given by the participants. Participants who presented their results in units other than those requested were asked to make the unit conversion themselves.

Table 4. Participant Results together with the Expanded Uncertainties and the Derived Consensus Values.

	Moisture Content			Free Fatty Acid Content		
Lab number[a]	**Result[b] ppm[d]**	**Uncertainty ppm**	**k[c]**	**Result[b] %**	**Uncertainty %**	**k[c]**
1	*217*			0.058	0.01	
2				0.0706	0.018	2; 95%
3	*300*			0.06		
4	280	18	2; 95%	0.043	0.005	2; 95%
5				0.042	0.006	2; 95%
6				0.032	0.01	2; 95%
7				0.073	0.002	2; 95%
8	249	18				
9	250	60	2; 95%	0.040	0.010	2; 95%
10				0.04	0.0022	95%
11				0.064	0.006	2; 95%
12				0.039	0.001	2; 95%
13	236.63	13.49	2; 95%			
14				0.057	0.053	95%
15						
Consensus value	**253.9**			**0.052**		

Lab number[a]	Peroxide value Result[b] meqO2/kg	Uncertainty meqO2/kg	k[c]	Phosphorus Content in crude rapeseed oil Result[b] ppm[d]	Uncertainty ppm	k[c]
1	2.69	0.12				
2	1.87					
3	3.04			145.18		
4	2.98	0.25	2; 95%	154	17	2; 95%
5	4.13	1.03	2; 95%			
6	4.92	1.0	2; 95%			
7	3.34			52.7[e]		
8						
9	2.72	0.30	2; 95%			
10	3.0	0.09	95%			
11	2.0	0.2	2; 95%			
12	3.9	0.39	2; 95%			
13						
14						
15				165.8		
Consensus value	**2.31**			**155.0**		

Lab number[a]	Saponification value Result[b] mg/g	Uncertainty mg/g	k[c]	Beta-Sitosterol content Result[b] ppm[d]	Uncertainty ppm	k[c]
1	198	3				
2	195.12					
3	190.57					
4	189.4	3.8	2; 95%			
5						
6						
7	190.5			3532.9		
8						
9				3300	300	2; 95%
10						
11	192	1	2; 95%	3600	180	2; 95%
12	188	2.1	2; 95%			
13						
14						
15						
Consensus value	**191.9**			**3477.6**		

	Erucic acid content		
Lab number[a]	Result %	Uncertainty %	k[c]
1			
2			
3	0.21		
4			
5	0.24	0.04	2; 95%
6			
7	0.229	0.018	2; 95%
8			
9			
10			
11	0.21	0.02	2; 95%
12	0.221	0.009	2; 95%
13			
14			
15			
Consensus value	**0.222**		

[a] The participating laboratories are given numbers in random order that is different from the order given in Table 3. [b] Moisture content values in italic were determined by the heating loss method and were not taken into account when determining the consensus value and target standard deviation. [c] Coverage factor or confidence level, as provided by the participants. [d] Although ppm is not an SI unit, it was decided to present the results in this unit respecting the established practice. ppm is identical to mg/kg. [e] Outlying result is marked in red.

The only result that was eliminated as outlier was the value of phosphorus 52.7 ppm. The reason was twofold. First of all, this low level of phosphorus content in hot-pressed crude rapeseed oil is technologically impossible and secondly, there are three results obtained using two principally different techniques (UV-Vis spectrophotometry of ashed oil and X-ray fluorescence analysis of oil without pretreatment), which agree among themselves and are roughly three times different from the eliminated result.[12]

Peroxide value (PV) is a special parameter because of its instability and in the past rounds of this intercomparison increasing of PV in time has often been observed. As Figure 4 demonstrates, there is no temporal trend in the PV values obtained by different laboratories at different times in this intercomparison round. The trend seems decreasing but the decrease is statistically insignificant. The better stability of the peroxide value is probably due to the antioxidant added to oil (vitamin e) in this round and possibly also to better sealing of the bottles.

According to our (SB) experiments (during the time period 25.05.2013-29.09.2013) PV increases in a closed bottle kept at room temperature by ca 0.05 meq/kg per ten days. Given that in the worst cases the measurements were done within a 40-day time window the reasonable increase of PV over time would be in the range of 0.2 meq/kg. It is possible that with some bottles the increase was faster, but in any case, the rate of PV increase is too vaguely defined to allow any meaningful correction. Therefore the all the PV values of the participants were used as they were reported, without applying any corrections.

Full information about consensus values and target standard deviations of all parameters is presented in Table 5 (the respective data of the EstOil rounds 1-7 are also given for reference).

[12] The outlying laboratory used UV-Vis spectrophotometry of ashed oil for phosphorus determination. The most difficult part of this procedure is dry-ashing of the oil. If not done extremely carefully, losses by splashing are very likely. This is a possible cause of the very low result.

Table 5. Consensus Values and Target Standard Deviations of Interlaboratory Comparison Measurements EstOil rounds 1-8.

Moisture content ppm	FFA content %	PV meq/kg	P content ppm	SAPV mg/g	B-SITO content ppm	EA content %	Unit
			EstOil-8[a]				
253.9	0.052	2.31	155.0	191.9	3477.6	0.222	consensus value
249.0	0.057	2.99	145.2	190.6	3532.9	0.222	consensus value (median)
20.9	0.016	1.85	11.7	3.8	178.5	0.015	target standard deviation
18.0	0.008	0.21	11.6	1.1	88.2	0.0	target standard deviation (median)
1.1	0.001	0.05	1.7	1.7	8.4	-[13]	between-sample variability[b]
8%	**30%**	**80%**	**8%**	**2%**	**5%**	**7%**	relative target standard deviation
7%	14%	7%	8%	1%	2%	5%	relative target standard deviation (median)
0.4%	1.6%	2.3%	1.1%	0.9%	0.2%	-	relative between-sample variability
1.9%	10.6%	29.7%	6.3%	0.7%	1.6%	0.2%	relative difference between two consensus value
			EstOil-7[a]				
155.7	0.044	1.31	280.6	189.1	3783.0	0.119	consensus value
151.7	0.038	1.29	260.5	189.0	3826.0	0.119	consensus value (median)
46.1	0.017	0.45	83.7	1.6	181.2	0.019	target standard deviation
21.5	0.005	0.15	31.1	1.2	119.4	0.0	target standard deviation (median)
2.6	0.001	0.0	5.1	0.0	148.1	0.0	between-sample variability[b]
30%	**38%**	**34%**	30%	0.9%	5%	16%	relative target standard deviation
14%	13%	11%	12%	0.6%	3%	10%	relative target standard deviation (median)
1.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	relative between-sample variability
2.6%	13.3%	1.8%	7.2%	0.1%	1.1%	0.0%	relative difference between two consensus value
			EstOil-6[a]				
367.0	0.041	2.57	112.7	188.7	3734.0	0.408	consensus value
357.0	0.037	1.63	112.3	189.0	3729.9	0.408	consensus value (median)
31.2	0.018	1.80	4.2	2.0	80.6	0.029	target standard deviation
16.1	0.004	0.21	1.1	0.8	85.3	0.0	target standard deviation (median)
1.4	0.0006	0.0	1.1	0.0	58.7	0.021	between-sample variability[b]
8%	**43%**	**70%**	4%	1.1%	2%	7%	relative target standard deviation
5%	11%	13%	1%	0.4%	2%	3%	relative target standard deviation (median)
0.4%	1.3%	0.0%	1.0%	0.0%	1.6%	5.1%	relative between-sample variability
2.7%	10.8%	36.5%	0.3%	0.2%	0.1%	0.0%	relative difference between two consensus value
			EstOil-5				
391.8	0.037	2.27	137.0	187.9	3920.5	0.282	consensus value
391.4	0.033	2.40	138.4	187.9	3913.5	0.284	consensus value (median)
6.3	0.014	0.82	7.4	3.8	50.7	0.019	target standard deviation
5.4	0.004	0.38	5.0	2.4	43.2	0.012	target standard deviation (median)
0.0	0.0005	0.05	1.0	1.3	101.2	0.003	between-sample variability[b]
2%	37%	36%	5%	2.0%	1%	7%	relative target standard deviation
1%	13%	16%	4%	1.3%	1%	4%	relative target standard deviation (median)
0.0%	1.3%	2.2%	0.8%	0.7%	2.6%	1.2%	relative between-sample variability
0.1%	10.1%	5.8%	1.0%	0.0%	0.2%	0.7%	relative difference between two consensus values

[13] For EA value the between-sample standard deviation could not be calculated because only two samples were used and the ANOVA results indicated that the overall variability is wholly due to within-sample variability.

EstOil-4							
373.7	0.036	1.71	130.5	190.2	3532.3	0.097	consensus value
365.0	0.033	1.51	123.0	191.0	3688.6	0.099	consensus value (median)
20.6	0.017	0.79	18.9	2.9	427.5	0.007	target standard deviation
13.0	0.004	0.18	8.4	0.7	132.9	0.005	target standard deviation (median)
0.5	0.002	0.00	2.2	1.0	27.0	0.006	between-sample variability[b]
6%	48%	46%	15%	1.5%	12%	7%	relative target standard deviation
4%	12%	12%	7%	0.4%	4%	5%	relative target standard deviation (median)
0.1%	4.2%	0.0%	1.7%	0.5%	0.8%	6.2%	relative between-sample variability
2.3%	8.5%	12.1%	5.7%	0.4%	4.4%	1.7%	relative difference between two consensus values

EstOil-3						
81.6	0.031	14.07	125.3	189.7	3877.4	consensus value
101.2	0.029	14.00	138.1	190.1	3820.5	consensus value (median)
22.4	0.013	3.60	55.4	2.3	423.5	target standard deviation
16.7	0.003	0.45	13.6	1.0	143.7	target standard deviation (median)
2.9	0.0000	0.20	2.7	1.0	59.1	between-sample variability[b]
27%	42%	26%	44%	1.2%	11%	relative target standard deviation
16%	11%	3%	10%	0.5%	4%	relative target standard deviation (median)
3.6%	0.0%	1.4%	2.2%	0.5%	1.5%	relative between-sample variability
23.9%	7.7%	0.5%	10.2%	0.2%	1.5%	relative difference between two consensus values

EstOil-2				
87.7	0.03	6.9	124.2	consensus value
19.1	0.01	1.9	31.9	target standard deviation
6.4	0.0003	0.03	1.9	between-sample variability[b]
22%	37%	27%	26%	relative target standard deviation
7.3%	1.1%	0.4%	1.5%	relative between-sample variability

EstOil-1		
362.5	0.07	consensus value
32.9	0.01	target standard deviation
6.0	0.0004	between-sample variability[b]
9%	19%	relative target standard deviation
1.7%	0.5%	relative between-sample variability

[a] The consensus values and target standard deviations were found using the Algorithm A described in the ISO 13528:2005 standard [b] Given at standard deviation level. See section 4.2 for more information. [c] For phosphorus content and EA value the between-sample standard deviation could not be calculated because the ANOVA results indicated that the overall variability is wholly due to within-sample variability. This indicates that between-sample variability is very low compared to the repeatability of the method itself. See section 4.2 for more information.

The participant z-scores are calculated according to equation 1 and are presented in Table 6.

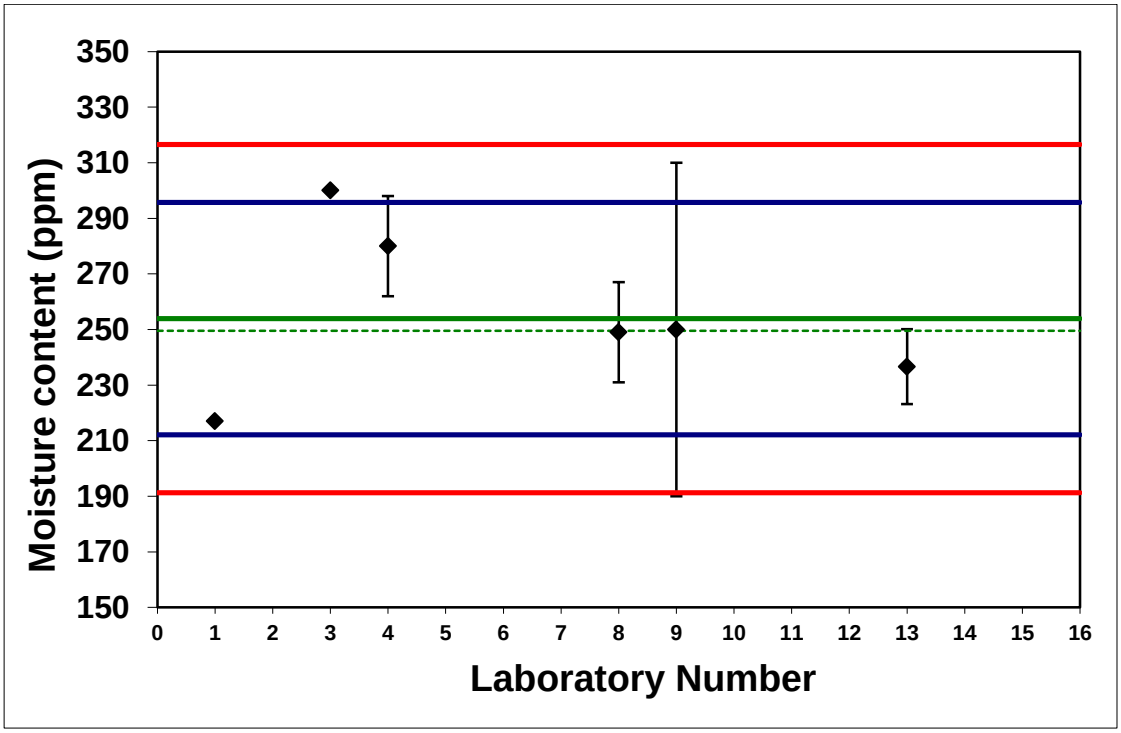
Table 6. Participant z-Scores found using the algorithm A.

	z scores[algorithm A]						
Lab number[a]	Moisture content	FFA	PV	P content	SAPV	B-SITO content	ERUC content
1	-1.8	0.4	0.2		1.6		
2		1.2	-0.2		0.8		
3	2.2	0.5	0.4	-0.8	-0.3		-0.8
4	1.2	-0.5	0.4	-0.1	-0.6		
5		-0.6	1.0				1.2
6		-1.3	1.4				
7		1.4	0.6	-8.7	-0.4	0.3	0.5
8	-0.2						
9	-0.2	-0.7	0.2			-1.0	
10		-0.7	0.4				
11		0.8	-0.2		0.03	0.7	-0.8
12		-0.8	0.9		-1.0		-0.1
13	-0.8						
14		0.3					
15				0.9			

[a] The participating laboratories are given in random order that is different from the order given in Table 3 but is identical to the order given in Table 4. [b] According to the ISO Guide 43-1: acceptable result is marked in green, doubtful result in yellow and unacceptable result in red.

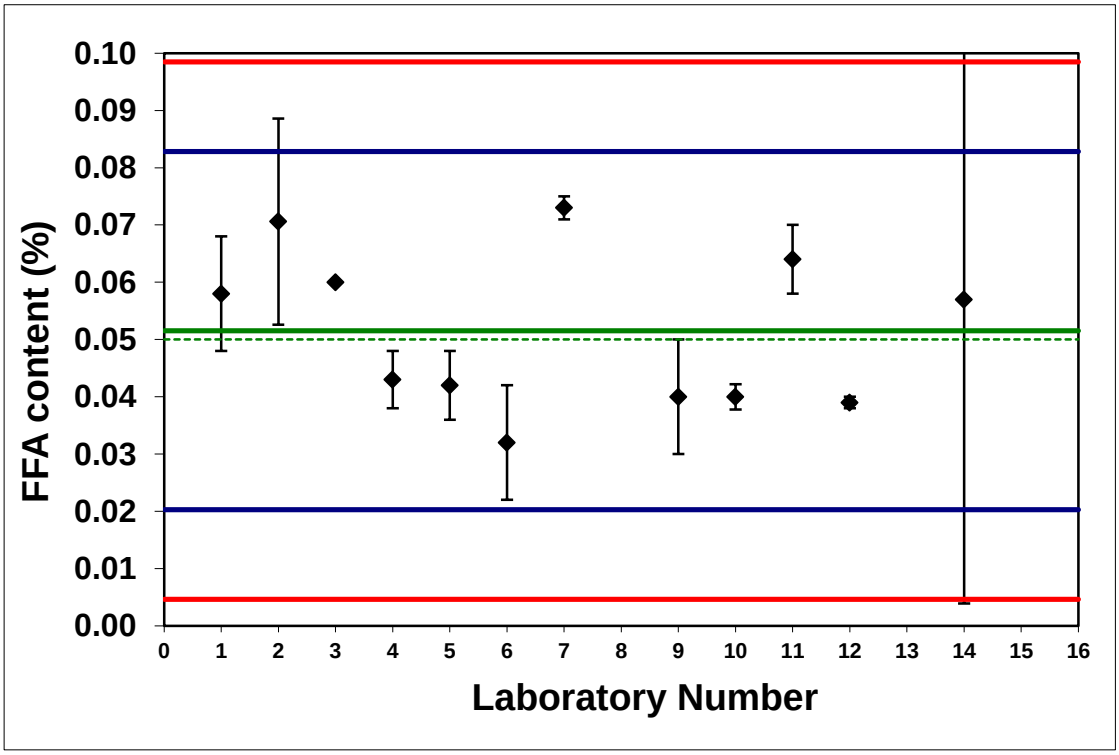
The results are presented in graphical form in the Figures below:

Figure 1. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] Moisture Content Measurement[b].



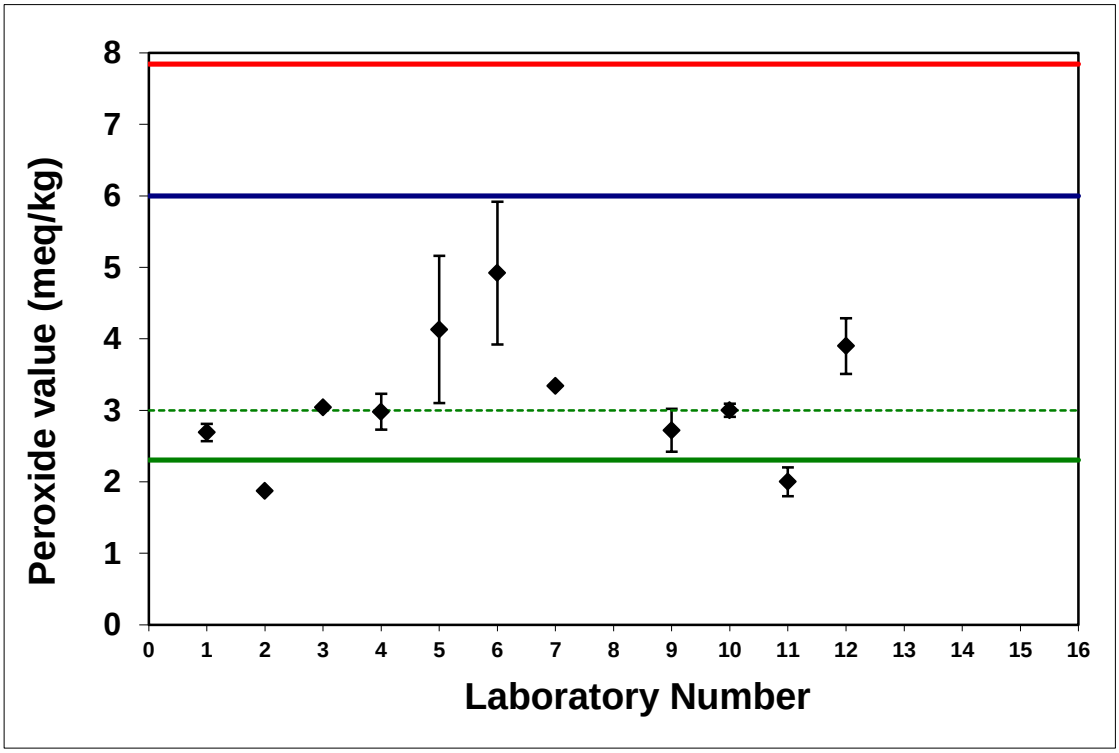
[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b]

Figure 2. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] FFA Content Measurement[b].



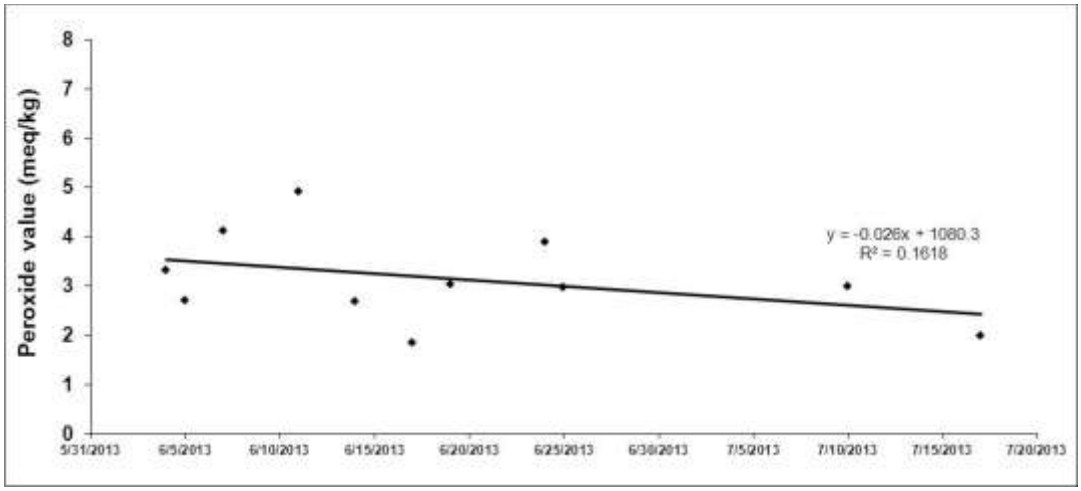
[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b]

Figure 3. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] PV Measurement[b].



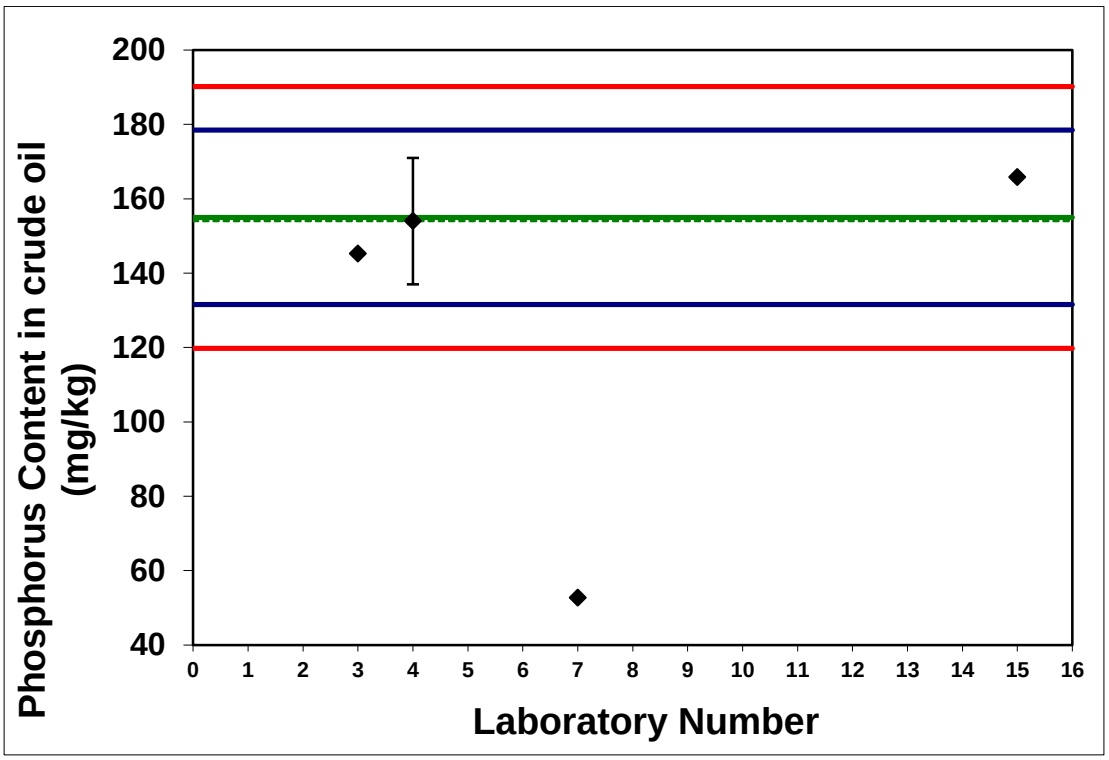
[a] The consensus value found using the algorithm A is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines. [b]

Figure 4. Dependence of the Results of PV Measurement on Measurement Date.[a]



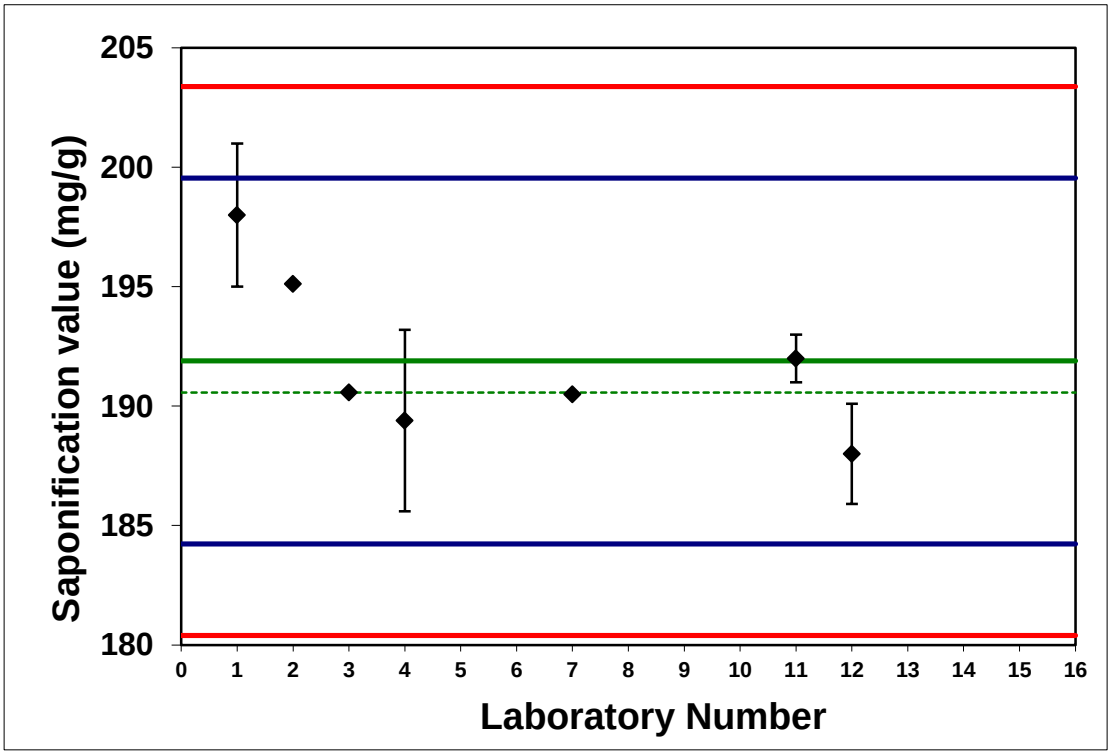
[a] back dots symbols denote measurements made by participants.

Figure 5. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] **P Content Measurement.**



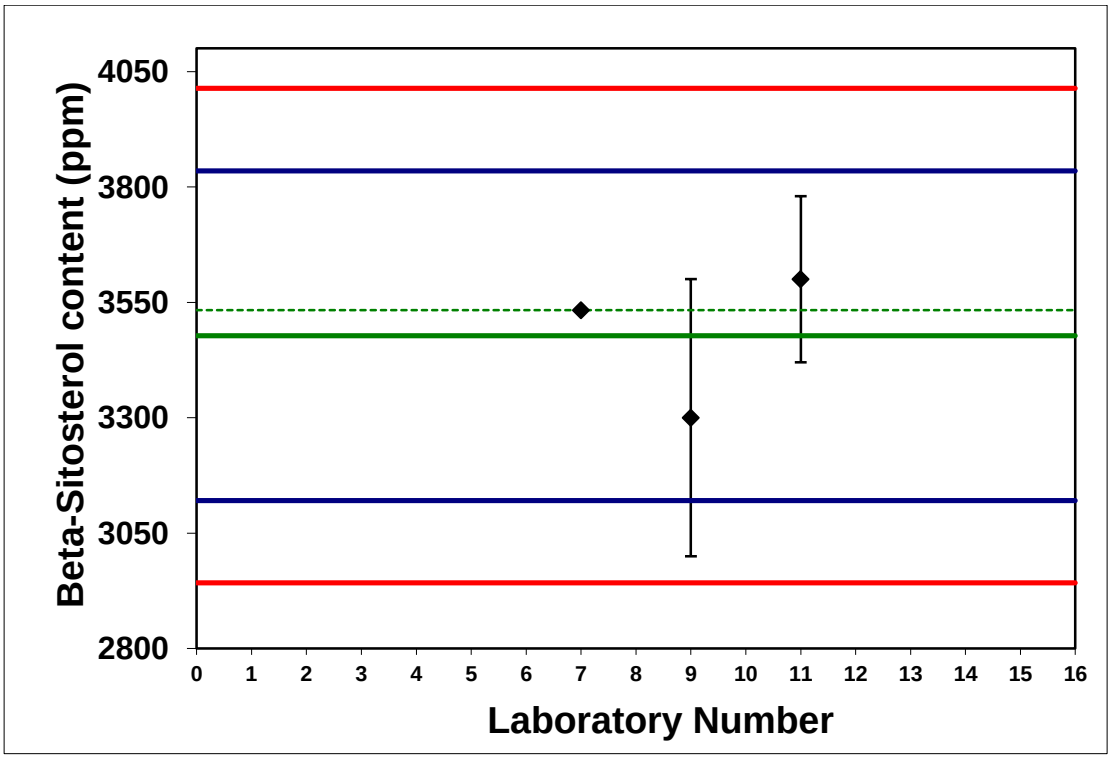
[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines.

Figure 6. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] SAPV Measurement.



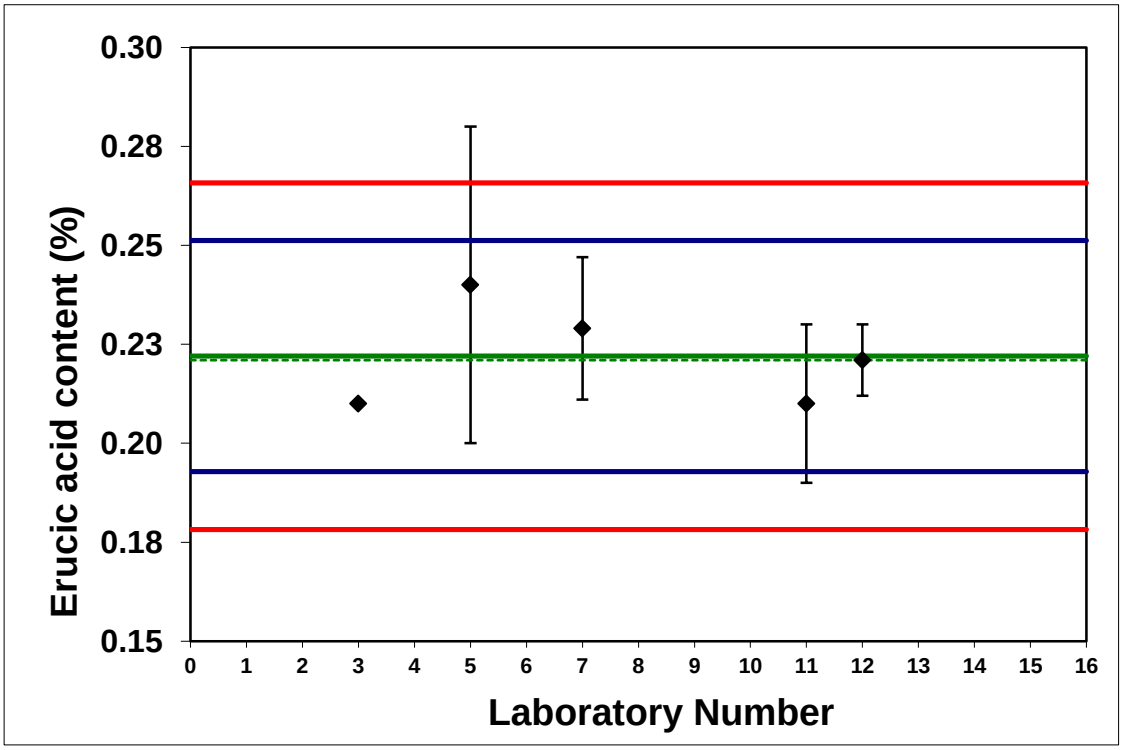
[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines.

Figure 7. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] B-SITO Measurement.



[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines.

Figure 8. Results of Participants with Self-declared Uncertainties and z-Score Boundaries.[a] EA Content Measurement.



[a] The consensus value found as average is denoted by the solid green line. The dotted green line denotes the median consensus value. The 2*s* boundaries are denoted by blue lines. The 3*s* boundaries are denoted by red lines.

From Table 6 and the Figures it can be concluded that based on the z-score approach from the 48 submitted results one is unacceptable and one is doubtful.

4.2 Between-Sample Variability

Between-sample variability was determined by UT (moisture content), by SB (FFA content, PV, P content, SAPV), by Institute of Chemistry, Tartu Veterinary and Food Laboratory (B-SITO content) and by Estonian Agricultural Research Centre (EA content) under repeatability conditions (see Table 5). The data were treated using the ANOVA technique to separate the effects of between- and within-sample variability.[14]

For moisture content the between-sample standard deviation was 1.1 ppm (three samples). This is 19 times lower than the between-participant standard deviation.

For FFA content the between-sample standard deviation was 0.001% (four samples). This is 19 times lower than the between-participant standard deviation.

For PV value the between-sample standard deviation was 0.05 meq/kg (three samples). This is 35 times lower than the between-participant standard deviation. For SAPV the between-sample standard deviation was 1.7 mg/g (three samples). This is 2 times lower than the between-participant standard deviation. Therefore with this parameter the between-sample variability is taken into account in the E_n score calculations (eq 3, see section 2.3).

[14] The treatment was carried out as described in A.M.H. van der Veen, J. Pauwels, *Accred. Qual. Assur.*, **2000**, *5*, 464-469.

For posphorus content the between-sample standard deviation was 1.7 ppm (three samples). This is 7 times lower than the between-participant standard deviation.

For B-SITO the between-sample standard deviation was found 8.4 ppm (two samples). This is 21 times lower than the between-participant standard deviation. The very large difference from the between-participant standard deviation partly compensates for the use of only two samples.

For EA value the between-sample standard deviation could not be calculated because only two samples were used and the ANOVA results indicated that the overall variability is wholly due to within-sample variability.

We can conclude that in moisture content, FFA content, PV, P content, B-SITO and EA content measurement the between-sample variability has negligible effect on the between-participant variability. In the case of SAPV measurement the between-sample variability is not negligible. Therefore with this parameter the between-sample variability is taken into account in the E_n score calculations (eq 3, see section 2.3).

5 Discussion

5.1 Assessment of Participant Results by the z-Score Approach

All in all 48 results were submitted. According to the z-score approach 46 of them (96%) were acceptable, one (2%) was doubtful and one (2%) was unacceptable (See Table 6).
As has been concluded in the previous rounds of this intercomparison with several measurands the large number of acceptable z-scores is caused by using the actual standard deviations as target standard deviations in z-score calculation. Standard deviations of FFA and PV content are 30% and 80% of the consensus value, respectively. The situation with moisture, SAPV, P, B-SITO and EA content (relative target standard deviation being 2% … 8%) is good or normal.

The very low spread of the participant values in SAPV determination also explains why the between-sample variability in these determinations is not lower than the target standard deviation. Based on the between-sample variability determinations it can be stated that in all other cases the spread of the participant results is primarily caused by the large between-lab scatter of the results.

FFA and PV deserve further comment here:

1. The FFA content in the samples was low. This is probably one of the main reasons for the large spread of the FFA measurement results. Several factors that at higher analyte levels are of low significance can seriously influence results at low analyte levels. In the case of e.g. FFA content these can be

    a. too small size of sample taken for titration and correspondingly low titrant volume, leading to the high sensitivity of the result towards end-point determination;[15]

    b. in the case of a larger sample, too low solvent volume, so that part of the sample may remain undissolved;[16]

    The content of FFA was similar to the previous EstOil rounds and the relative target standard deviations are also similar, indicating consistent performance of the participants.

2. PV is a very unstable parameter.

[15] The end point determination, if phenolphtalein is used as indicator, is not trivial in FFA determination, especially at low levels. The end point is not stable – the rose color disappears with time. It needs skill and experience to determine for how long time the color has to persist in order to claim it a true end point (generally around 15 s).

[16] In FFA determination it is important to add a sufficient amount of solvent so that all sample dissolves. The SB laboratory used the mixture of isopropyl alcohol and diethyl ether.

The high spread of the FFA results is caused first of all by the low FFA content of the samples.

The high spread of the PV results is caused first of all by the instability of this parameter.

Although according to the z-score approach most of the participants performed satisfactorily, it is of interest to compare the self-declared uncertainties of the participant results to the agreement between the participant results. The picture is not encouraging. B-SITO, EA and possibly moisture content are the only measurands where the uncertainties have in general been estimated realistically.

In the case of all measurands except B-SITO, EA and moisture a significant number of participants have severely underestimated their uncertainties.

In addition, many participants still report results without any uncertainty estimate.

Obviously, especially when having data with high spread, the z-score approach has serious deficiencies when assessing participant performance:

(1) Uncertainties of participant results are not taken into account.

(2) The consensus values derived from the participant data are too unreliable to be used as reference values.

(3) The standard deviations of the participant data are too large and result in excessively wide acceptable result zones. Because of the low reliability of the consensus value it is also not reasonable to use a narrower predefined target standard deviation, because then it is possible that laboratories obtaining correct results will have unacceptable z-scores.

These problems would be solved by independently determined reference values for the samples. However, this would make the intercomparison significantly more expensive. The main conclusion is:

In spite of the generally good z-scores, the participants should carefully and critically examine their results and their agreement with other laboratories!

5.2 Assessment of Participant Results by the Robust Approach

The median-based robust data treatment approach is less sensitive to outliers than those obtained with the classical z-score approach. Therefore both the consensus value and the target standard deviation are more "robust" (see Annex 1 for more information). As an important consequence – the target standard deviations are lower and thus the assessment is stricter. Results of data treatment using this approach are presented in Annex 1.

One can see that 11 results are found unacceptable with this approach, compared to one result found doubtful using the z-score approach.[17]

[17] The robust approach does not distinguish between doubtful and unacceptable results

Even though the results presented in Annex 1 are not official and are of informative nature, we encourage the participants whose results are unacceptable according to Annex 1 carefully examine their analysis procedures and uncertainty budgets.

5.3 Pairwise Comparison of Participant Results

Under these circumstances a useful alternative is the pair-wise comparison of laboratory results using the E_n scores. This approach is not as informative as comparison against an independent reference value but is still useful – it issues a clear warning signal to laboratories whose results disagree from the results of most other laboratories.

The paired comparisons are presented in Tables 7 to 12.

Table 7. Comparison of the Results of Participants in Pairs: Moisture Content Measurement.

	$\lvert E_n \rvert$ value			
Lab No	4	8	9	13
4		1.2	0.5	1.9
8	1.2		0.02	0.5
9	0.5	0.02		0.2
13	1.9	0.5	0.2	

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: $\lvert E_n \rvert \leq 1$ is considered acceptable agreement between two results. Acceptable agreement is marked in green and unacceptable in yellow.

Table 8. Comparison of the Results of Participants in Pairs: FFA Measurement.

	$\lvert E_n \rvert$ value										
Lab No	1	2	4	5	6	7	9	10	11	12	14
1		0.6	1.3	1.4	1.8	1.5	1.3	1.8	0.5	1.9	0.0
2	0.6		1.5	1.5	1.9	0.1	1.5	1.7	0.3	1.8	0.2
4	1.3	1.5		0.1	1.0	5.6	0.3	0.5	2.7	0.8	0.3
5	1.4	1.5	0.1		0.9	4.9	0.2	0.3	2.6	0.5	0.3
6	1.8	1.9	1.0	0.9		4.0	0.6	0.8	2.7	0.7	0.5
7	1.5	0.1	5.6	4.9	4.0		3.2	11.1	1.4	15.2	0.3
9	1.3	1.5	0.3	0.2	0.6	3.2		0.0	2.1	0.1	0.3
10	1.8	1.7	0.5	0.3	0.8	11.1	0.0		3.8	0.4	0.3
11	0.5	0.3	2.7	2.6	2.7	1.4	2.1	3.8		4.1	0.1
12	1.9	1.8	0.8	0.5	0.7	15.2	0.1	0.4	4.1		0.3
14	0.0	0.2	0.3	0.3	0.5	0.3	0.3	0.3	0.1	0.3	

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: $\lvert E_n \rvert \leq 1$ is considered acceptable agreement between two results. Acceptable agreement is marked in green and unacceptable in yellow.

Table 9. Comparison of the Results of Participants in Pairs: PV Measurement.

	$\lvert E_n \rvert$ value							
Lab No	1	4	5	6	9	10	11	12
1		1.05	1.4	2.2	0.1	2.1	3.0	3.0
4	1.05		1.1	1.9	0.7	0.1	3.1	2.0
5	1.4	1.1		0.6	1.3	1.1	2.0	0.2
6	2.2	1.9	0.6		2.1	1.9	2.9	0.95
9	0.1	0.7	1.3	2.1		0.9	2.0	2.4
10	2.1	0.1	1.1	1.9	0.9		4.6	2.2
11	3.0	3.1	2.0	2.9	2.0	4.6		4.3
12	3.0	2.0	0.2	0.95	2.4	2.2	4.3	

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: $\lvert E_n \rvert \leq 1$ is considered acceptable agreement between two results. Acceptable agreement is marked in green and unacceptable in yellow.

In the case of P Content Measurement this table cannot be compiled because there is only one laboratory that has evaluated measurement uncertainty. However, inspection of the data clearly reveals that the agreement between three laboratories is good, while the result of the fourth laboratory strongly disagrees.

Table 10. Comparison of the Results of Participants in Pairs: SAPV Measurement.

	$\lvert E_n \rvert$ value			
Lab No	1	4	11	12
1		0.98	0.8	1.2
4	0.98		0.3	0.2
11	0.8	0.3		0.5
12	1.2	0.2	0.5	

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: $\lvert E_n \rvert \leq 1$ is considered acceptable agreement between two results. Acceptable agreement is marked in green and unacceptable in yellow. Because the between sample variability uncertainty of SAPV measurement is relatively high, it was necessary to use the modified equation (eq 3) for E_n value calculation.

Table 11. Comparison of the Results of Participants in Pairs: B-SITO content Measurement.

	$\lvert E_n \rvert$ value	
Lab No	9	11
9		0.9
11	0.9	

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: $\lvert E_n \rvert \leq 1$ is considered acceptable agreement between two results. Acceptable agreement is marked in green and unacceptable in yellow.

Table 12. Comparison of the Results of Participants in Pairs: EA Content Measurement.

	$\|E_n\|$ value			
Lab No	**5**	**7**	**11**	**12**
5		0.3	0.7	0.5
7	0.3		0.7	0.4
11	0.7	0.7		0.5
12	0.5	0.4	0.5	

[a] The numbers of participants are the same as in Table 4. [b] According to the ISO Guide 43-1: $|E_n| \leq 1$ is considered acceptable agreement between two results. Acceptable agreement is marked in green and unacceptable in yellow.

From Tables 7 to 12 it can be seen that in determination of SAPV, B-SITO and EA content determination between-lab agreements dominate. Disagreeing comparisons dominate in PV (75%) measurement. At the same time the z-scores of the participants in general look good. This clearly indicates, that most of the participants should take a close look at their uncertainty estimates of these measurements. The abovementioned factors provide some guidelines. As a conclusion:

> **The pair-wise agreement of participant results in SAPV, B-SITO and EA content determination is good to satisfactory while in the determination of moisture content, FFA and PV it is unsatisfactory.**

6 Acknowledgments

The organizers would like to thank Mr Viktor Vabson (AS Metrosert, Division of Standards Services), Dr Olev Saks (Testing Centre, University of Tartu), Dr. Koit Herodes (Testing Centre, University of Tartu), Mrs Vaimi Nõmm (Tartu Veterinary and Food Laboratory), Mrs Ann Akk (Estonian Agricultural Research Centre) for advice and help.

7 Annex 1

The usual statistical algorithm of finding z scores[9] may give not the best estimates of z scores of the participants with several of the analytes determined in this intercomparison. The reasons for this are: (1) there are serious outliers (gross errors) among the data, (2) the results of the determinations carried out near the detection limit cannot be assumed to be normally distributed and (3) the results of different participants cannot be assumed to have the same uncertainty. Therefore the arithmetic mean may not be the ideal consensus value and z score may not be the ideal performance criterion.

Below we apply an alternative data analysis procedure based on the E_n scores using median and its uncertainty as the estimate of the consensus value and its uncertainty, respectively.[11] Arithmetic mean value is known to lack "robustness" – stability against outliers. Median has a significantly better statistical robustness.

For a continuous variate C_{lab}, the median as a consensus value C_c is defined, using the (cumulative) distribution function $F(C_{lab})$, by the condition:

$$F(C_c) = \frac{1}{2} \ . \tag{4}$$

This means that one half of the observations are below and the other above the median. For sample of n ordered variables C_{lab1}, C_{lab2}, …, C_{labn}, the sample median, denoted as C_c=med{C_{labi}}, is given by (with integer k)

$$C_c = \begin{cases} C_{labk+1}, \ldots\ldots\ldots\ldots\ldots k = \dfrac{n-1}{2} \ for\ odd\ n \\ \tfrac{1}{2}(C_{labk} + C_{labk+1}), \ldots\ldots k = \dfrac{n}{2} \ for\ even\ n \end{cases} \tag{5}$$

Uncertainty of median is found as follows:

$$u(C_c) = D \cdot MAD \tag{6}$$

where D is defined as follows:

$$D = \frac{1.858}{\sqrt{n-1}} \tag{7}$$

and the where the value MAD is given by:

$$MAD = med\left\{\left|C_{labi} - C_c\right|\right\}, for\ \ i= 1, 2, \ldots, n. \tag{8}$$

The median-based consensus values for the measurands are given in Table 5.

Assessment of the results is done using the E_n numbers as described in ISO Guide 43-1:[9]

$$\left|E_n\right| = \frac{\left|C_{lab} - C_c\right|}{\sqrt{U_{lab}^2 + U_c^2}} \tag{9}$$

where C_{lab} are the results of a laboratory, C_c is the median as a consensus value and U_{lab} and U_c are the expanded uncertainties of the laboratory value and the median, respectively. Equation 9 is adequate, if between-sample variability is significantly (more than 5 times) lower than between-

participant variability. If not, the between-sample variability has to be taken into account and the E_n value is found as follows:

$$|E_n| = \frac{|C_{lab} - C_c|}{\sqrt{U_{lab}^2 + U_c^2 + (t_{95}(df) \cdot s_s)^2}} \quad (10)$$

where s_s – is the between-sample standard deviation and $t_{95}(df)$ is the student coefficient at 95% confidence level with df degrees of freedom.

Agreement between two results is considered acceptable if $|E_n| \leq 1$.

The results of this data treatment are presented in Table 13.

Table 13. Participant $|E_n|$ values according to the robust approach.

	$\|E_n\|$ scores[b] accdording to median approach						
Lab number[a]	Moisture content	FFA	PV	P content	SAPV	B-SITO content	EA content
1	0.9	0.5	0.7		0.9		
2		0.97	2.8		0.6		
3	1.4	0.9	0.1	0.5	0.0		0.5
4	0.8	0.6	0.0	0.0	0.1		
5		0.6	1.02				0.4
6		1.2	1.8				
7		2.0	0.9	5.4	0.0	0.0	0.3
8	0.0						
9	0.0	0.7	0.6			0.7	
10		0.9	0.0				
11		1.1	2.2		0.2	0.3	0.4
12		0.98	1.6		0.3		0.0
13	0.3						
14		0.1					
15				0.6			

[a] The participating laboratories are given in random order that is different from the order given in Table 3 but is identical to the order given in Table 4. [b] According to the ISO Guide 43-1: acceptable result is marked in green and unacceptable result in yellow. The results of the participants who did not report uncertainties were assigned zero uncertainty.

The results presented in Table 13 are of informative nature for the current intercomparison round but more investigations will be performed and in the future this data treatment may be considered as the definitive one. If this decision will be made then this will be stated in the invitation to the intercomparison.